*Proc. Natl. Acad. Sci. USA*
Vol. 88, pp. 3295–3299, April 1991
Evolution

# Retrotransposable elements *R1* and *R2* interrupt the rRNA genes of most insects

(transposable elements/sequence specificity/molecular evolution)

JOHN L. JAKUBCZAK, WILLIAM D. BURKE, AND THOMAS H. EICKBUSH

Department of Biology, University of Rochester, Rochester, NY 14627

*Communicated by Igor B. Dawid, January 2, 1991*

ABSTRACT **A large number of insect species have been screened for the presence of the retrotransposable elements *R1* and *R2*. These elements integrate independently at specific sites in the 28S rRNA genes. Genomic blots indicated that 43 of 47 insect species from nine orders contained insertions, ranging in frequency from a few percent to >50% of the 28S genes. Sequence analysis of these insertions from 8 species revealed 22 elements, 21 of which corresponded to *R1* or *R2* elements. Surprisingly, many species appeared to contain highly divergent copies of *R1* and *R2* elements. For example, a parasitic wasp contained at least four families of *R1* elements; the Japanese beetle contained at least five families of *R2* elements. The presence of these retrotransposable elements throughout Insecta and the observation that single species can harbor divergent families within its rRNA-encoding DNA loci present interesting questions concerning the age of these elements and the possibility of cross-species transfer.**

Transposable elements probably exist in the genome of every eukaryote (1). Among the most abundant types of these elements are the retrotransposable elements that, like retroviruses, move by means of an RNA intermediate (2). Retrotransposable elements can be divided into two major classes (3, 4). One class is similar to the retroviruses in that they contain long terminal repeats (LTRs) and their encoded proteins have amino acid similarity to those of the retroviruses. The second class, termed Line 1-like or the non-LTR retrotransposable elements, lacks any type of terminal repeat and has lower levels of amino acid similarity to the retroviruses.

*R1* and *R2* (formerly called type I and type II insertions) are non-LTR retrotransposable elements, each found at a precise location in a fraction of the 28S rRNA genes of *Bombyx mori* and several dipteran species (5–12). The insertion sites for *R1* and *R2* are 74 base pairs (bp) apart in a highly conserved region of the 28S gene, the large subunit rRNA gene of eukaryotes (Fig. 1*A*). The presence of either *R1* or *R2* within an rRNA-encoding DNA (rDNA) unit inactivates that unit (13–15). Only a few copies of *R1* and *R2* are located outside the rDNA units and they appear to be nonfunctional (6, 16, 17). The high insertion specificity of *R2* elements can be explained by an encoded endonuclease activity specific to its 28S gene insertion site (18). Individual copies of *R1* and *R2* have the same 3′ end but may be truncated at their 5′ end, a feature common to other non-LTR retrotransposable elements (6, 7, 11, 17).

Although transposable elements as a group are widespread, it has been difficult to study the distribution of a particular element across broad taxonomic groups due to their dispersed genomic locations and rapid sequence divergence that limits their detection by DNA hybridization. Here we have taken advantage of the remarkable insertion specificities of *R1* and *R2* to determine their distribution in species throughout the class Insecta. We present evidence that *R1* and *R2* are present in the rRNA genes of most insects. Unexpectedly, we found frequent examples of single insect species harboring multiple, highly divergent families of *R1* or *R2* elements within their rDNA loci.

## MATERIALS AND METHODS

**Insect Species.** *Forficula auricularia, Libellula pulchella, Mantis religiosa, Popillia japonica, Malacosoma americanum, Tibicen* sp., *Sphecius speciosus, Dissoteira carolina, Leptinotarsa decemlineata,* the Xylocopinae species, and the Carabidae species were collected from the wild. *Acheta domestica, Blaberus craniifer, Hippodamia convergens,* and *Oncopeltus fasciatus* were purchased from Carolina Biological Supplies. Other species were obtained from laboratory-maintained stocks.

**Genomic Blots.** For each insect, 1- to 3-μg samples of total adult DNA was digested with one or more of the following restriction enzymes: *Hin*cII, *Hin*cII plus *Eco*RI, *Hin*cII plus *Hin*dIII, *Msp* I, *Hae* III, *Sau*3A, and *Taq* I. Genomic blotting and hybridization were performed as described (12) except that the hybridization was done at 60°C without calf thymus DNA. Blots were probed with a *Bam*HI–*Hin*dIII fragment from the 28S gene of *Drosophila melanogaster,* clone a56 (12). This fragment contained 280 bp of 28S gene and 18 bp of *R1* sequence. The percentage of the 28S genes with insertions in each species was calculated by laser densitometric scanning of autoradiographs.

**Cloning and DNA Sequencing.** A Charon 35 library of parasitic wasp (*Nasonia vitripennis*) DNA (19) and an EMBL 4 library constructed of Japanese beetle DNA were hybridized with the same probe used in the genomic blots. *Eco*RI–*Hin*cII restriction fragments corresponding to variant 28S genes were subcloned into M13mp18 and M13mp19 vectors (20). To obtain the 3′ junction of the elements with the 28S gene, the mp19 clones were sequenced using Sequenase (United States Biochemical) and the polynucleotide 5′-ACGGTCTGATCTCAGTTCGA-3′ to prime synthesis at a site 65 bp downstream of the *R1* insertion site. Nucleotide sequences of regions within the insertion elements were obtained from the mp18 clones using the "Universal" primer supplied by United States Biochemical. The 3′ junctions of the cockroach and praying mantis elements were obtained by cloning 0.9- to 1.2-kilobase-pair (kb) *Eco*RI–*Hin*cII fragments from a genomic digest directly into mp19. The 3′ junction sequences of the 28S gene insertions from the cicada, cicada killer, grasshopper (*D. carolina*), and carpenter bee were amplified by the inverse polymerase chain reaction (21) of *Msp* I- or *Mbo* I-digested genomic DNA using

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: rDNA, rRNA-encoding DNA; 28S gene, large subunit rRNA gene of eukaryotes; LTR, long terminal repeat; ORF, open reading frame.

the primers 5′-CCATGAGCTCGGTTTCGCTAGATAGTAGATAGGGAC-3′ and 5′-CGATGAATTCCCTGTTGAGCTTGACTCTAGTCTGGC-3′. These primers (containing *Eco*RI and *Sst* I tails) are complementary to conserved regions of the 28S gene downstream of the *R1* site.

## RESULTS

**Genomic Blot Assay.** The presence of *R1* and *R2* elements in different insect species was detected using the assay diagramed in Fig. 1*A*. Genomic DNA from each species was digested with a series of restriction enzymes, blotted onto nitrocellulose paper, and probed with a 280-bp fragment of the *D. melanogaster* 28S gene immediately 3′ of the *R1* and *R2* insertion sites. Because the ribosomal gene sequence is highly conserved within a species, uninterrupted 28S genes generate a single hybridizing restriction fragment, whereas 28S genes containing insertions in the region of the *R1* and *R2* target sites give rise to fragments of different molecular weights.

Examples of genomic blots for 10 insect species are shown in Fig. 1*B*. In each case, the DNA was digested with *Hin*cII alone or in combination with a second restriction enzyme. *Hin*cII cleaves ≈60 bp 5′ of the *R2* site in most insects; thus

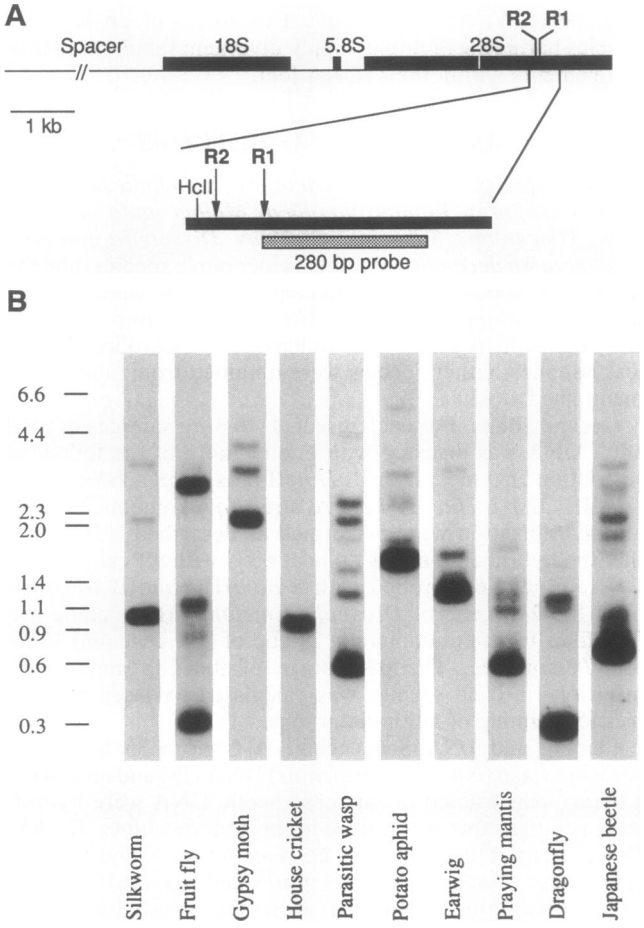


FIG. 1. Procedure used to identify insertion elements within a region of the 28S rRNA genes of insects. (*A*) Schematic diagram of a typical rDNA repeat unit in insects. The black boxes represent the regions that correspond to the mature 18S, 5.8S, and 28S rRNAs; the line represents the transcribed and nontranscribed spacer regions. Locations of the *R1* and *R2* insertion sites are indicated. The *Hin*cII site (HcII) 5′ of the *R2* insertion site is conserved among insects examined in this study. The shaded box represents the 280-bp 28S gene sequence used as hybridization probe. (*B*) Genomic blots of DNA from 10 insects hybridized with the 28S gene probe. The lowest molecular weight band in each lane corresponds to the uninterrupted 28S genes. The less intense higher molecular weight bands correspond to insertions in the 28S gene between the *Hin*cII site 5′ of the *R2* insertion site and either an *Eco*RI or *Hin*cII site 0.3–2.0 kb downstream of the *R2* insertion site. Sizes are indicated in kb. Exposure times were adjusted so that the intensity of the band corresponding to the uninterrupted genes was approximately the same in each lane. Genus and species name for each insect can be found in Table 1. Silkworm, *B. mori*; fruit fly, *D. melanogaster*.

Table 1. Summary of insects screened for insertions near the *R1* and *R2* sites

| Order/species | Common name | 28S genes inserted, % |
|---|---|---|
| Odonata | | |
| *L. pulchella* | Dragonfly | 10 |
| Orthoptera | | |
| *Melanoplus femurrubrum* | Grasshopper | ≤5 |
| *D. carolina* | Grasshopper | ≤5 |
| *A. domestica* | House cricket | 0 |
| *M. religiosa* | Praying mantis | 35 |
| *B. craniifer* | Giant cockroach | 25 |
| *Locusta migratoria* | African migratory locust | 20 |
| *Tibicen* sp. | Cicada | 10 |
| Dermaptera | | |
| *F. auricularia* | Earwig | 10 |
| Hemiptera | | |
| *O. fasciatus* | Milkweed bug | ≤5 |
| *Rhodnius prolixus* | Assassin bug | ≤5 |
| Homoptera | | |
| *Macrosiphum euphorbiae* | Potato aphid | ≤5 |
| *Pseudococcus affinis* | Mealybug | 0 |
| Coleoptera | | |
| *Tribolium confusum* | Confused flour beetle | 15 |
| *H. convergens* | Ladybug | ≤5 |
| *Tenebrio molitor* | Yellow mealworm | 20 |
| *P. japonica* | Japanese beetle | 30 |
| *L. decemlineata* | Colorado potato beetle | 40 |
| Family: Carabidae | Ground beetle | 0 |
| Hymenoptera | | |
| *N. vitripennis* | Parasitic wasp | 25 |
| *Nasonia giraulti* | Parasitic wasp | 15 |
| *Nasonia longicornis* | Parasitic wasp | 30 |
| *Trichomolopsis dubius* | Parasitic wasp | 15 |
| *S. speciosus* | Cicada killer | 30 |
| Subfamily: Xylocopinae | Carpenter bee | 15 |
| *Atta cephalotes* | Leafcutter ant | ≤5 |
| Lepidoptera | | |
| *B. mori* | Silkmoth | 10–30 |
| *Bombyx mandarina* | Silkmoth | 25 |
| *Manduca sexta* | Tobacco hornworm | 20 |
| *Lymantria dispar* | Gypsy moth | 30 |
| *M. americanum* | Eastern tent caterpillar | 0 |
| Diptera | | |
| *Drosophila* (12 species)* | Fruit fly | <5–65 |
| *Glossina pallidipes* | Tsetse fly | ≤5 |
| *Aedes aegypti* | Yellow fever mosquito | 10 |
| *Culex quinquefasciatus* | Southern house mosquito | 10 |
| *Sarcophaga bullata* | Flesh fly | ≤5 |

The percentage of the 28S genes with insertions in each species was calculated by laser densitometric scanning of autoradiographs similar to those in Fig. 1*B*. Because of the error involved in comparing bands of very different intensities, all values have been rounded off to the nearest 5%. Species were scored as having 0% of their 28S genes inserted if they did not have variant bands representing at least 1% of the total 28S genes, reproducibly present with different restriction enzymes.

*Species screened in this report were *melanogaster, mauritiana, simulans, pseudoobscura, immigrans, falleni, quinaria, subpaulstris, testecea, putrida, virilis,* and *busckii.*

the lowest molecular weight band in each lane represents the uninterrupted 28S genes, whereas the higher molecular weight bands represent 28S genes with insertions in or near the *R1* and *R2* target sites. The first two lanes in Fig. 1*B* are positive controls. In the silkmoth, *B. mori*, the two faint upper bands represent equivalent levels of the *R1* and *R2* elements (17). In the fruit fly, *D. melanogaster*, the upper intense band represents *R1* elements, the 1.1-kb band represents lower levels of *R2* elements, and the 0.8-kb band represents 5′ truncated copies of *R1* that are 0.5 kb in length (22). Of the eight additional insect species shown in Fig. 1*B*, only the house cricket does not show evidence of 28S gene insertions.

One unexpected finding from our screen was the number of variant bands detected in many species. Although some of the insects had a relatively simple pattern of one or two variant bands (e.g., gypsy moth and earwig), most insects tested had a complex pattern of variant bands (e.g., parasitic wasp, potato aphid, praying mantis, and Japanese beetle). These "variant" bands do not represent partial digestions or restriction site polymorphisms within the 28S gene sequence since variant bands reproducibly appeared when the DNA from each species was digested multiple times with different restriction enzymes.

A total of 47 species from nine orders of insects was tested by this genomic blot method (Table 1). Forty-three of these species were found to have insertions in a fraction of their 28S genes. The four species in which inserts were not detected (house cricket, mealybug, ground beetle, and Eastern tent caterpillar) are from different orders of insects, indicating that species without insertions are not limited to a particular taxonomic group. Densitometer scans of the autoradiographs derived from these genomic blots revealed that the percentage of the rDNA units containing insertions varied from <5% to >50%. Because the data in Table 1 are based upon a single individual or a small number of individuals from each species, they do not reflect the range of insertion levels that may exist in each species. For example, a 3- to 5-fold range in percentage of rDNA units inserted is found among different strains of *B. mori* and *D. melanogaster* (ref. 17; J.L.J. and T.H.E., unpublished data).

***R1* and *R2* Junction Sequences.** To determine whether the variant bands detected in the genomic blots corresponded to insertions in the *R1* and *R2* sites, eight species were selected for sequence analysis. For each species, DNA segments representing variant bands seen on the genomic blots were either cut directly out of agarose gels, obtained by inverse polymerase chain reaction (21), or isolated from genomic libraries (see *Materials and Methods*). The nucleotide sequences of the 3′ junction between the insertion elements and the 28S gene are shown in Fig. 2, and their relative positions within the 28S gene are shown in Fig. 4.

Insertion elements were found in the *R1* site in all eight species (Fig. 2*A*) and insertions in the *R2* site were found in three species (Fig. 2*B*). Unlike the situation in *B. mori* and several dipteran species (5–12), considerable sequence variation was detected among the 28S gene insertions within some species. For example, seven clones of parasitic wasp 28S genes containing insertions in the *R1* site were obtained. Based upon nucleotide sequence comparisons of a 150-bp

| | | R1 Elements | 28S gene |
|---|---|---|---|
| **A** | | | |
| Parasitic wasp | Nv 18,23 | GCCGTGTGGGT-GAGC--ACAAGCAGGAACCGCACGTTAAATCATCTAGGGCTTCTCAA | **TGTCCCTATCTA** |
| | Nv 1,12 | .....AA.TA...CGGA.G.....TGT.....A.G..----A..... | ..... |
| | Nv 8 | .....A.A.....CTCT....AGC.....-...A..A--C.----GACA.. | ..... |
| | Nv 4,16 | .....G..CG.A.G.....T..T.....A..AAG..----A.GATC | ..... |
| Cicada killer | Ss 1 | GACTTGGGAACGGCGACGGTCCGATTCATTTTGGATCGAGACATATCGCACATAATACT | **TGTCCCTATCTA** |
| | Ss 3,7 | .....--C.....C..... | ..... |
| | Ss 13 | .....--C.....C.A.....G..... | ..... |
| | Ss 14 | .....--C.....C.-.....T..... | ..... |
| | Ss 8 | T.TCA.TAC..TCG.GTAC.GG.GAGGCCC.AT.CCGCTCGACGGCGCGCG.C.C..A. | ..... |
| Cicada | Ti 2,5,6 | AGGTGGGCC-GGTGAAAAGCGAAAGCTTACATGCTGGCGGAGAGGGTTTGGAGACTAAC | **TGTCCCTATCTA** |
| | Ti 1 | C..CC..GTT.....TG.....T..A.CA.TCCGTCATCCGATT...GG... | ..... |
| | Ti 4 | .....-.....T.T...TA | ..... |
| Carpenter bee | Xy 1 | ATACTCCAGGAATGGCACGGACGGCCCCAAAGCCGAGCC-CTTGGACGCTCAAACTACT | **TGTCCCTATCTA** |
| | Xy 2 | .....G...T.....-.....A...AT.....T....A. | ..... |
| Grasshopper | Dc 1 | TGGCAGATGGCAACATCGCCGGTGTGGTAAAACCAAATCGGCGAGCCCATGTACGGTAC | **TGTCCCTATCTA** |
| Cockroach | Bc 2 | /ACCACGCCCTGACTTAGAGCCCTAGAAAGTCAGGACACGAGTATTGCAGGTTCGGT$_{38}$ | **TGTCCCTATCTA** |
| Praying mantis | Mr 2,4,5 | GACCCGGGTCTCGTTCCTACGGGGACGGGGCTAATGGGAGTAGACTCTCTGCACTCAGA | **TGTCCCTATCTA** |
| Japanese beetle | Pj 3 | AGCCGTATGTTTTGCAGGTCGCGTGCCGCCCCTTCATAGGGGAGCAGCGAGTTGTTTTC | **TGTCCCTATCTA** |
| **B** | | **R2 Elements** | |
| Japanese beetle | Pj 1 | ACGGCGGGTTACCACCTAATGCCGAAATGTACTTCATAAAATGAAATAAAGTATTATAA | **TAGCCAAATGCC** |
| | Pj 10,16 | ...AAC.CCA.TGTG.AC.ATG.TTTG.T.CG.ATTATTTGAT.....-..TA.. | ..... |
| | Pj 10a | .A.TGCAACATG.TTTAT..TTAT.TT.TATT.AT..T..TGC.....-.ATAC. | ..... |
| | Pj 12 | ..AA.CC.AA.TGTGT..TATATAT.TATACTA.GTATT.TATC.....CA-..TA.. | ..... |
| | Pj 19 | TTC....CAA..AT.AG.GCCAG...GGT..TAAATGTTTT.ATT....T.G.GAGA.C | ..... |
| Praying mantis | Mr 1 | GACTCCTTAAAATCCGAATTGGGGGGCACTCCAGTTTTTTCGGGGGCCTCTTAGTCACT | **TAGCCAAATGCC** |
| | Mr 6 | T.A.T.A.TT.G.TTA.CCCAACA.TGGGGTACAC.GGAAAT..TTTTATACTAGG.TA | ..... |
| Cicada killer | Ss 4,6 | TGTAGCATAAGTGGGAGATGGTAACATCGCCGGTGTCGCAAGGCACTAGGTATAGATTT | **TAGCCAAATGCC** |
| **C** | | **Other** | |
| Japanese beetle | Pj 17 | AAGTCTGCTGGCGGCCGGTCGGCGGTTCCTGCGCTCGTCCGCTTCGGCCCCTCGGCTCG | **GCGGTAGTAACT** |

FIG. 2. The 3′ junction sequences of insertion elements with the 28S gene. The 28S gene sequences are in boldface. When multiple *R1* or *R2* elements were obtained from the same species their nucleotide sequences were compared. A period (.) indicates identical nucleotides; a dash (−) indicates addition of nucleotides in one or more elements. A single slash, in the cockroach insertion, represents the 5′ end of the presumably 5′ truncated insertion. The 3′ end of this element contains a tract of 38 Ts. Grasshopper, *D. carolina*; parasitic wasp, *N. vitripennis*.

region at the 3′ junctions, these clones could be divided into four families. Nucleotide sequence identity between these families varied from 49% to 78%. The sizes of the *Eco*RI–*Hin*cII fragments generated by these families corresponded to the four major variant bands detected on the genomic blot shown in Fig. 1*A*. In addition, internal sequences from two clones (Nv1, Nv8) were used to probe genomic blots to demonstrate that most copies of these two families of elements are located within the 28S genes and that each element corresponded to a single variant band in the genomic blots (data not shown). In the case of the Japanese beetle, six of the eight clones isolated contained insertions in the *R2* site. Based upon the nucleotide sequence at their 3′ ends, these could be divided into five families with little, if any, nucleotide similarity. The cicada killer and cicada were each found to have at least two families of elements inserted in the *R1* site, and the praying mantis had two families of insertion elements in the *R2* site. This is not a complete listing of elements present in these eight species since only a small number of clones were analyzed for each species, and in most cases these clones do not represent all of the bands detected on genomic blots.

In only one instance was a 28S gene insertion found that was not in either the *R1* or *R2* site. In the Japanese beetle an insertion was found 27 bp upstream of the *R2* site. Genomic blots probed with an internal segment of this insertion (data not shown) indicated that most copies of this element are in the 28S genes, giving rise to the 1.9-kb variant band detected in Fig. 1*B*. Further analysis is necessary to determine whether this insertion represents another retrotransposable element specific to the 28S gene.

***R1* and *R2* Internal Sequences.** A parasitic wasp and the Japanese beetle were selected for further analysis since they had complex patterns of variant genomic bands and genomic libraries were available. The nucleotide sequence of internal segments was determined for three families of inserts in the *R1* site (Nv8, Nv18, and Pj3). Conceptual translation revealed an open reading frame (ORF) in each element with amino acid similarity to the putative integrase region of the *R1* elements of *B. mori* and *D. melanogaster* (12). Most revealing was the presence of the conserved cysteine motif shown in Fig. 3*A*. Although a similarly located cysteine motif is found in many non-LTR retrotransposable elements (12), the exact spacing of the cysteine and histidine residues as well as the presence of other conserved residues clearly identify the wasp and beetle insertions as *R1* elements. The level of amino acid identity between the two wasp elements was only 47%, indicating that they represent evolutionarily distinct *R1* elements. Further analysis will reveal whether the remaining wasp *R1* families are evolutionarily distinct.

Conceptual translation of internal sequences obtained from three families of insertions in the *R2* site of the Japanese beetle (Pj1, Pj10, and Pj12) revealed an ORF with significant amino acid similarity to the putative integrase region of the *R2* elements (12). This *R2* integrase region, like that in *R1* elements, contains a distinctive cysteine motif that can be identified in these beetle elements (Fig. 3*B*). The level of amino acid identity between the ORFs of the three sequenced beetle R2 elements varied from 44% to 55%.

**A R1 Elements**

| | |
|---|---|
| Silkmoth | CDCGAVEEDRD**H**VLWEC |
| Fruit fly | CACGDPYEDWM**H**ILCAC |
| Parasitic wasp (Nv8) | CACGTEREDWI**H**VLCEC |
| " (Nv18) | CMCGDVSEDWR**H**VLCRC |
| Japanese beetle (Pj3) | CECGEAEETAD**H**VWWEC |
| | * ** * * * |

**B R2 Elements**

| | |
|---|---|
| Silkmoth | CRAGCKVRETTA**H**ILQQC |
| Fruit fly | CRAGCDAPETTN**H**IMQKC |
| Japanese beetle (Pj1) | CRAGCERVESVS**H**ILQAC |
| " (Pj10) | CRGGCGKQATIS**H**VLQRC |
| " (Pj12) | CRNGCPRTESLS**H**VLQGC |
| | ** ** * * * |

FIG. 3. Comparison of the cysteine motif in the integrase region of *R1* (*A*) and *R2* (*B*) elements. Sequences for the silkmoth (*B. mori*) and fruit fly (*D. melanogaster*) elements are from ref. 12. Conserved cysteine and histidine residues are shown in boldface. *, Positions conserved in all examples.

## DISCUSSION

**Most Insects Contain *R1* and/or *R2* Elements.** The genomic blot assay described in this report indicated that 43 of 47 insect species contained insertions near the *R1* and *R2* sites in a fraction of their 28S rRNA genes. Sequence analysis of the insertions from eight species confirmed that all but one of the insertions are precisely located at the *R1* or *R2* sites. In two species, internal sequences within the ORFs of the insertions reveal amino acid sequence similarity to the *R1* and *R2* elements of *D. melanogaster* and *B. mori*. The analysis presented here was an attempt to sample the insertions detected in different species to determine whether these insertions corresponded to *R1* and *R2* rather than an attempt to detect all 28S gene insertions within particular species. Based on this analysis we conclude that most of the variant bands detected on the genomic blots of insect species correspond to *R1* or *R2* elements.

The only insertion detected in this study that was not located in the *R1* or *R2* site, Pj17 from the Japanese beetle, is an insertion 27 bp upstream of the *R2* site. Three other 28S gene insertions have been previously reported (see Fig. 4). A 28S gene insertion, named *R3*, which generated a 14-bp deletion in the same region of the 28S gene as the Pj17 insertion, has been detected in the fungus gnat, *S. coprophila* (11). An insertion between the *R1* and *R2* sites, which generates a target site duplication from 7 to 14 bp in length, has been detected in the nematode, *A. lumbricoides* (24). Finally, an insertion that generates a 17-bp duplication has been detected ≈600 bp downstream of the *R1* site in the mosquito, *Anopheles gambiae* (25). For all three elements most copies are located in the 28S genes, they do not contain terminal repeats, and 5′ truncated copies are known to exist. Sequence analysis should reveal whether these elements represent independent non-LTR retrotransposable elements that became specialized for insertion into the 28S gene or elements related to *R1* and *R2* that have changed their insertion specificity.

**Evolution of *R1* and *R2*.** The broad distribution for *R1* and *R2* found in this study indicates that they have been highly successful at adapting to survival and propagation within the rDNA locus. By comparing the phylogeny of these elements from various species with the phylogeny of their insect hosts, it should be possible to resolve the issue of whether these transposable elements were present in insects before their radiation and stably maintained for >300 million years or whether they are capable of being horizontally transferred between species. Indeed, our analysis has already revealed at least one surprising finding. *R1* or *R2* elements within a given species can be members of highly divergent families. The parasitic wasp was found to have at least four families of *R1* elements and Japanese beetle was found to have at least five families of *R2*. The level of amino acid identity between these different families can be quite low (50–60%), approaching that seen in comparisons between *B. mori* elements and *D. melanogaster* elements (12). The presence of multiple families of *R1* or *R2* in the same species clearly differs from the

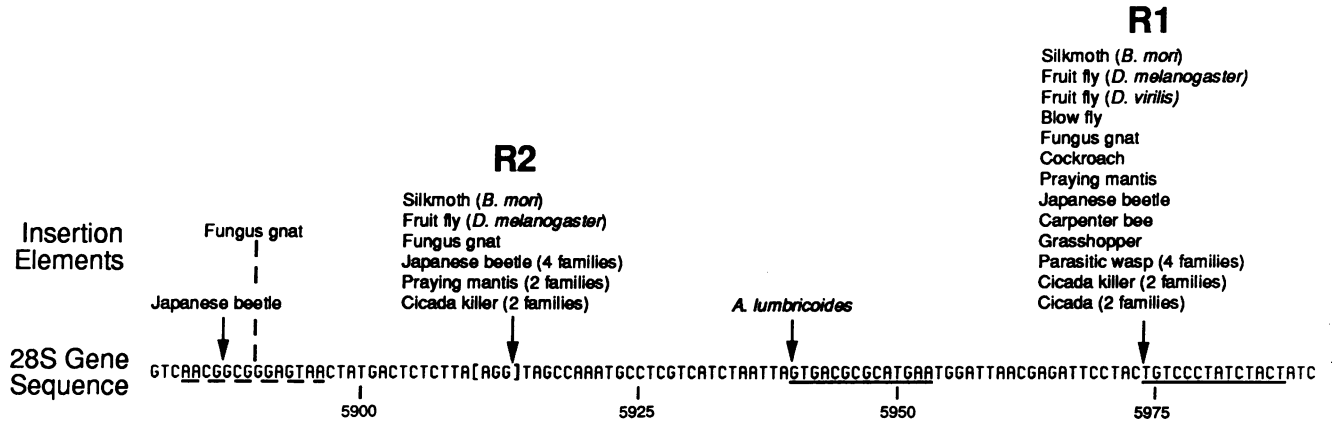


FIG. 4. The 28S gene insertions in or near the *R1* and *R2* sites. The sequence of the uninterrupted 28S gene of *D. melanogaster* is shown across the bottom and is numbered as in Tautz *et al.* (23). Locations of the insertions isolated from each organism are indicated. Data were obtained from the references indicated: *D. virilis* (5); blowfly, *Calliphora erythrocephala* (8); fungus gnat, *Sciara coprophila* (11); *Ascaris lumbricoides* (23). Nucleotides that are frequently deleted from the 28S gene upon the insertion of R2 are shown in brackets. The 14 bp of 28S gene sequence duplicated upon the insertion of *R1* elements or the *A. lumbricoides* element are underlined with a solid line. Deleted nucleotides associated with the fungus gnat element are underlined with a dashed line. Because of this deletion the exact location of the original insertion is not known.

situation in the previously studied species where *R1* or *R2* elements from the same species are nearly identical in sequence (5–12).

The presence of such highly divergent families of *R1* or *R2* elements in the same species is difficult to explain. If these families have coexisted in the same species for the time required for this extensive divergence, then each family must have been able to regulate its copy number during this period in such a way as to prevent its own elimination and at the same time fail to displace the other families in the competition for a limited number of 28S gene insertion sites. An alternative explanation is that *R1* and *R2* elements can be transferred between species with the result that the number of families within a given species reflects the present state of a process of introduction of new families and elimination of families by competition or stochastic loss.

Finally, *R1* and *R2* are of interest for another reason. Several transposable elements have been successfully developed or are being actively pursued as transformation systems for the introduction of new DNA sequences into their hosts. The ability of *R1* and *R2* to propagate in a wide range of insects makes either of these elements an attractive candidate for the development of a transformation system that would be applicable to most insects. A considerable advantage in efforts to achieve this goal is the opportunity to develop the system first in *D. melanogaster*.

We thank the following individuals for supplying insects or insect DNA: S. Aksoy, B. Beard, P. Blader, L. Beukeboom, D. Cox-Foster, M. Eickbush, A. James, J. Jaenike, C. Mullen, C. Nichols-Orians, U. Nur, B. Sakaguchi, E. Stephenson, R. Stouthammer, K. Wallace, J. Werren, and G. Wyatt. We thank J. Jaenike, U. Nur, and M. Palopoli for comments on an earlier version of the manuscript. This work was supported by National Institutes of Health Grant GM31867 and American Cancer Society Grant NP-691.

1. Berg, D. E. & Howe, M. M. (1989) *Mobile DNA* (Am. Soc. Microbiol., Washington).
2. Boeke, J. D. & Corces, V. (1989) *Annu. Rev. Microbiol.* **43,** 403–434.
3. Xiong, Y. & Eickbush, T. H. (1988) *Mol. Biol. Evol.* **5,** 675–690.
4. Xiong, Y. & Eickbush, T. H. (1990) *EMBO. J.* **9,** 3353–3362.
5. Rae, P. M. M., Kohorn, B. D. & Wade, R. P. (1980) *Nucleic Acids Res.* **8,** 3491–3504.
6. Roiha, H., Miller, J. R., Woods, L. C. & Glover, D. M. (1981) *Nature (London)* **290,** 749–753.
7. Dawid, I. B. & Rebbert, M. L. (1981) *Nucleic Acids Res.* **9,** 5011–5020.
8. Smith, V. L. & Beckingham, K. (1984) *Nucleic Acids Res.* **12,** 1707–1724.
9. Burke, W. D., Calalang, C. C. & Eickbush, T. H. (1987) *Mol. Cell. Biol.* **7,** 2221–2230.
10. Xiong, X. & Eickbush, T. H. (1988) *Mol. Cell. Biol.* **8,** 114–123.
11. Kerrebrock, A. W., Srivastava, R. & Gerbi, S. A. (1989) *J. Mol. Biol.* **210,** 1–13.
12. Jakubczak, J. L., Xiong, Y. & Eickbush, T. H. (1990) *J. Mol. Biol.* **212,** 37–52.
13. Jamrich, M. & Miller, O. L., Jr. (1984) *EMBO J.* **3,** 1541–1545.
14. Long, E. O. & Dawid, I. B. (1979) *Cell* **18,** 1185–1196.
15. Kidd, S. J. & Glover, D. M. (1981) *J. Mol. Biol.* **151,** 645–662.
16. Browne, M. J., Read, C. A., Roiha, H. & Glover, D. M. (1984) *Nucleic Acids Res.* **12,** 9111–9122.
17. Xiong, Y., Burke, W. D., Jakubczak, J. L. & Eickbush, T. H. (1988) *Nucleic Acids Res.* **16,** 10561–10573.
18. Xiong, Y. & Eickbush, T. H. (1988) *Cell* **55,** 235–246.
19. Nur, U., Werren, J. H., Eickbush, D. G., Burke, W. D. & Eickbush, T. H. (1988) *Science* **240,** 512–514.
20. Yanish-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
21. Ochman, H., Gerber, A. S. & Hartl, D. L. (1988) *Genetics* **120,** 621–623.
22. Wellauer, P. K., Dawid, I. B. & Tartof, K. D. (1978) *Cell* **14,** 269–278.
23. Tautz, D., Hancock, J. M., Webb, D. A., Tautz, C. & Dover, G. A. (1988) *Mol. Biol. Evol.* **5,** 366–376.
24. Back, E., Van Meir, E., Muller, F., Schaller, H., Neuhaus, P., Aeby, P. & Tobler, H. (1984) *EMBO J.* **3,** 2523–2529.
25. Paskewitz, S. M. & Collins, F. H. (1989) *Nucleic Acids Res.* **17,** 8125–8133.